基于 CLV 偏好挖掘模型的数字社区用户偏好挖掘研究

肖　耘[1]，许欢欢[1]，肖雅元[1]，赵又霖[2,3*]，庞航远[3]

（1. 广西中烟工业有限责任公司，南宁　530001；
2. 南京大学 信息管理学院，南京　210023；3. 河海大学 商学院，南京　211100）

摘　要：［目的 / 意义］数字社区已经成为企业高效管理用户的一种方式，用户行为信息以及用户的客户生命周期价值对数字社区的用户偏好挖掘具有重要意义。且现有的数字社区研究缺乏对用户价值和未来偏好挖掘的研究。［方法 / 过程］针对数字社区的用户群体，本文提出基于客户生命周期价值 CLV（Customer Lifetime Value，CLV）的偏好挖掘模型 CLV-PM（CLV-Preference Mining，CLV-PM）。首先，为反映用户真实偏好，基于用户行为信息，借助 RFM 模型和 K-Means++ 算法挖掘用户群体特征，生成用户价值类别标签；其次，为考虑用户时序性和差异性以及增强模型对偏好的认知，利用用户 CLV 构建用户-评分矩阵，并借助协同过滤算法挖掘用户预测偏好；最后，绘制数字社区目标用户的用户偏好画像。［结果 / 结论］“微信社群”管理平台的用户数据集中，可划分为重要价值用户、低价值用户、回流用户和重要挽留用户 4 种用户价值类别；目标用户 16254 为重要价值用户，采取“留存和维持”为主的运营策略；历史偏好为欢乐跳一跳、秒杀等活动，预测偏好为飞行棋大作战、猜码图等活动，目标用户偏好画像为数字社区运营和维护用户提供依据。

关键词：CLV-PM；协同过滤；数字社区；用户偏好；信息行为

中图分类号：G250　　**文献标识码：**A　　**文章编号：**1002-1248（2023）02-0045-16

引用本文：肖耘，许欢欢，肖雅元，等. 基于 CLV 偏好挖掘模型的数字社区用户偏好挖掘研究[J]. 农业图书情报学报，2023，35(2)：45-60.

1 引　言

随着信息技术和互联网的迅猛发展，数字化社区应运而生并迅速发展。数字社区作为一种全新的生活方式，以数字技术为基础，通过网络、手机等终端进行信息传播和交流。然而，由于数字社区的用户信息

收稿日期：2023-01-10

基金项目：广西中烟工业有限责任公司科技项目“基于机器学习方法的营销活动效果动态评估”（CGAXZX20210030050001-044）；江苏省社会科学基金青年基金“社会感知数据驱动下的公共卫生事件时空演化研判机制研究”（20TQC001）；中国博士后科学基金特别资助“面向应急管理的时空数据语义模 型构建及创新应用机理研究”（2021T140311）；中国博士后科学基金面上项目“环境污染突发事件的时空数据挖掘及协同治理机制研究”（2019M650108）

作者简介：肖耘（1971-），硕士，研究方向为“互联网 +”营销产品产品研发、生产与运营。许欢欢（1988-），女，硕士，研究方向为互联网营销及研究。肖雅元（1988-），女，研究方向为互联网营销及研究。庞航远（2002-），女，硕士研究生，研究方向为知识组织研究

*通信作者：赵又霖（1986-），女，副教授，博士生导师，南京大学博士后，河海大学商学院，研究方向为数据分析与挖掘、知识组织研究。E-mail：20140068@hhu.edu.cn

飞速增长，信息过载问题相继出现，用户难以从海量数据资源中找到自身需要的物品。

数字社区用户生命周期描述了用户参与社区活动的不同阶段，不同的用户拥有不同的生命周期，并且用户对于社区的价值贡献和需求存在差异。因此，衡量数字社区用户的客户生命周期价值不仅考虑了用户的差异性，而且考虑了用户的时序性。

在众多的用户偏好挖掘研究算法中，协同过滤算法的应用最为普遍。协同过滤依赖于偏好或兴趣与目标用户相似的用户，并推荐用户可能感兴趣的项目。由于传统的协同过滤算法的实现非常依赖物品和用户的评分信息，但用户的评分信息往往伴随数据稀疏性和数据真实性问题，而用户行为信息能够真实反映用户的偏好，有效减少数据的稀疏性和失真性问题。

因此，为提高预测和挖掘的精度，考虑用户时序性和用户价值。本文将用户行为数据作为数据源，从客户生命周期价值的视角出发，构建 CLV-PM（CLV-Preference Mining，CLV-PM）模型。通过聚类划分用户价值类别，生成用户价值类别标签，挖掘用户历史偏好，再结合协同过滤算法预测用户未来偏好，最后，生成数字社区用户偏好画像，为数字社区用户的偏好挖掘提供依据。同时，为数字社区用户偏好挖掘提供新的研究视角。

2 相关研究基础

2.1 CLV 理论及应用

客户生命周期价值 CLV（Customer Lifetime Value，CLV）[1]是一项用于衡量客户贡献利润的典型指标，对企业的精准营销具有重要的价值和意义，其测量和计算被广泛应用于学术研究和营销领域。现有研究成果表明，以 CLV 为基础的营销资源分配为企业带来了更多的利润。VENKATESAN 和 KUMAR[2]发现前 5% 的顾客所创造的价值要比其他模型高出 10%~15%；KUMAR 等[3]指出 CLV 模型可以帮助企业衡量客户关系，制定更为合理的营销政策，实现个性化管理，使客户价值最大化；李玉婷等[4]指出 CLV 高的企业，其客户续保率越高、赔付率越低。有关 CLV 的主要研究内容可以分为：用户价值细分[4]、客户生命周期建模[5]、CLV 对相关决策管理的支持[6,7]等。

2.2 数字社区用户研究现状

数字社区是指通过数字信息将服务提供者和管理部门与用户连接起来的虚拟在线社区，而数字社区的用户是指使用由服务提供者提供的服务的人。近年来，数字在线社区方面的研究引起了学者们的广泛关注，并得到了许多出色研究成果。数字在线社区的研究主要涉及用户信息披露[8,9]、用户行为影响因素[10-18]、用户偏好挖掘[19-23]等。由于本文涉及数字社区用户偏好挖掘以及用户行为方面的分析，下面将重点阐述这两个方面的数字社区研究现状。

在用户行为分析方面，肖雪等[10]以“豆瓣读书”作为数据来源，通过社会网络分析法、内容分析法和统计分析法分析虚拟阅读社区的用户互动特征和影响因素；普哲缘和李胜利[11]以哔哩哔哩作为数据来源，借助双向固定效应模型探究视频评论特征对观众评论行为的影响；付少雄等[12]以好大夫在线作为数据来源，基于社会基本理论探究在线医疗社区医生知识贡献行为的关键动因；潘涛涛和吕英杰[15]以某在线健康社区的发帖行为数据为数据源，借助 SOA 模型探究影响用户参与社区意愿的因素；赵欣等[16]以问卷数据作为数据来源，运用 AMOS 软件探究用户行为与用户信任的互惠因果关系；陈星等[18]以问卷数据作为数据来源，运用 AMOS 探究影响用户持续知识分享行为意愿的因素。

在用户偏好挖掘方面，学者主要以用户评论数据、用户基本属性以及用户行为数据等为研究数据来源；借助扎根理论、标签分类、聚类分析以及情感分析等方法展开用户需求主题识别、关键用户识别等研究。如成全和郑抒琳[24]以母婴网站的提问数据作为数据源，分析其用户信息需求主题标签体系，并构建层级多标签分类模型；余佳琪等[8]基于患者的评论数据构建了一个挖掘不同阶段患者评论主题与情感状态的主题情感混合模型；吴江等[19]以网易云社区为研究对象，借助

BERT 主题聚类的方法，分析不同音乐主题的特征；张军等[20]从用户交互行为属性、信息质量属性和情感倾向属性 3 个方面展开关键用户识别研究；王帅[21]从用户的基本属性、兴趣主题、情感倾向、问诊需求以及社交网络 5 个方面进行用户画像和用户分群研究；钱宇星等[23]以“老年人之家”论坛中的文本为数据源，借助共现分析和主题分析挖掘老年在线健康社区的健康信息需求（表 1）。

2.3 协同过滤算法研究现状

协同过滤算法是目前推荐系统中应用范围最广且成功率最高的推荐算法。常常被应用于预测和挖掘用户的需求和偏好。传统的协同过滤算法通常基于用户对项目的评分数据预测用户偏好[25]。但是，评分信息的失真问题导致预测结果不够精确，因此学者们提出结合文本内容和社区网络[26-28]、用户的属性信息[29,30]、时空信息[31]、用户的浏览、复制以及收藏信息[32-34]等提高结果的准确性。

从 CLV 的理论及应用来看，CLV 作为用户价值衡量的重要指标，其对资源的有效利用和用户价值的最大化具有重要的地位，且被广泛应用于用户价值衡量领域，为基于用户价值的用户偏好挖掘提供新的视角；从数字社区用户的研究现状来看，数字社区的用户行为研究主要集中在用户行为的影响因素方面，数字社区用户偏好研究多以文本数据作为数据来源，少以用户的行为数据作为研究对象，而用户的行为数据真实反映用户的偏好；借助主题分析等方式挖掘用户的偏好，少有对未来偏好的预测研究；现有的数字社区用户研究少有考虑用户生命周期价值，但用户生命周期价值反映用户整个生命周期内对数字社区的贡献，考虑用户生命周期价值有利于挖掘和预测用户偏好，从而提高数字社区的运营效率和效果。从协同过滤算法的研究现状来看，单一评分数据存在失真问题，现有研究采用多属性特征结合的方法提高预测和挖掘的精确度。

综上所述，以现有的研究为基础，本文提出了一种以用户行为数据作为研究对象，考虑用户生命周期价值的混合用户偏好挖掘模型——CLV-PM 模型。该模型从 CLV 的视角出发，将用户行为数据作为用户偏好数据，评估和计量用户的 CLV；以用户的 CLV 为衡量指标，利用 K-means++ 算法进行用户聚类，生成用户价值类别标签，最后通过协同过滤算法挖掘不同用户价值类别的未来偏好，并在此基础上绘制数字社区中目标用户的用户偏好画像，为数字社区用户的偏好挖掘提高依据。

3 CLV-PM 模型的构建

为克服现有研究的局限性以及数字社区“信息过载”的问题，并基于数字社区用户的时序性以及用户价值差异性的特点。本文提出一种基于 CLV 的偏好挖掘模型——CLV-PM，用于数字社区的用户偏好挖掘研究。CLV-PM 模型的作用有二：一是提高偏好挖掘和预测的准确度，以用户行为数据作为研究对象，真实反映用户偏好；二是从用户的 CLV 的视角出发，进行用户聚类，生成用户价值类别标签，实现数字社区资源最大化，用户价值最大化。CLV-PM 模型的算法框架如图 1 所示。

3.1 RFM 模型

在数字社区中，用户的评分存在失真或数据稀疏性问题，RFM 模型通过量化用户行为信息，对用户进

表 1　数字社区用户偏好挖掘研究特征表

Table 1　Digital community user preference mining research

数据类型	用户评论数据、用户基本属性、用户行为数据等
研究方法	扎根理论、层级多标签分类、K-means 方法、EM 聚类、情感分析、BERT 主题聚类、AttriRank 算法、共现分析等
研究主题	用户需求主题识别、关键用户识别等

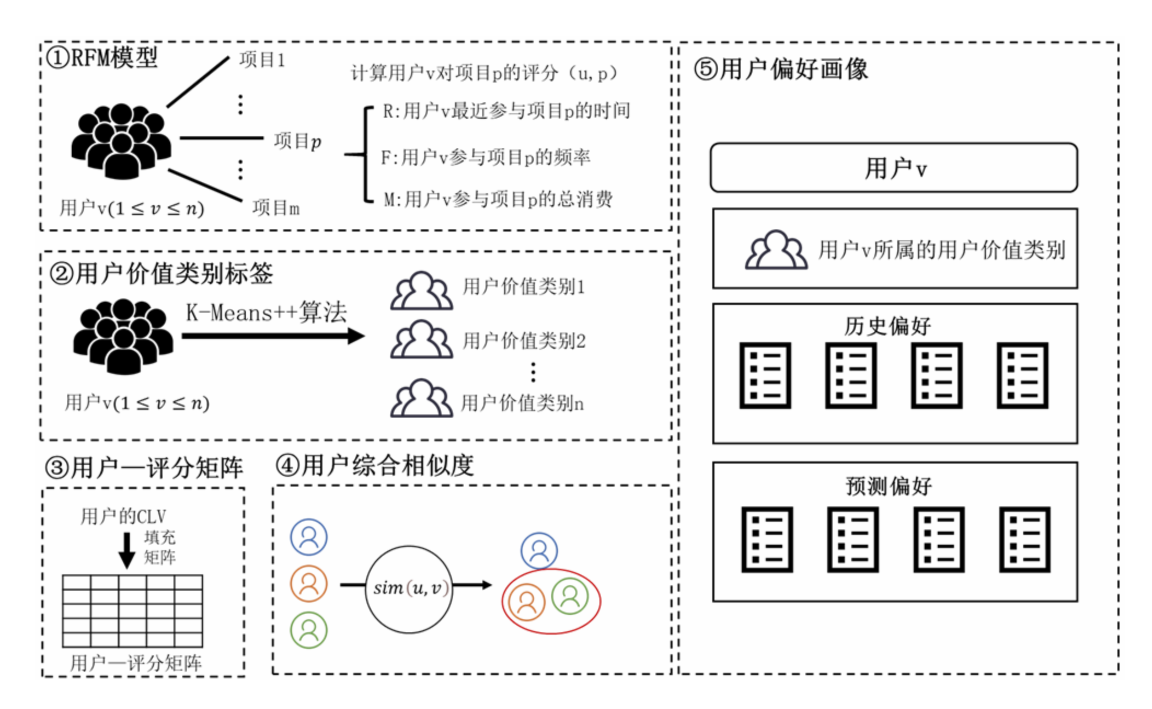


图 1　CLV-PM 模型计算框架

Fig.1 CLV-PM model calculation framework

行价值划分，以此衡量用户对社区的评分。本文通过基于 RFM 模型量化数字社区的用户行为信息，以挖掘用户偏好和衡量用户价值。RFM（Rational Function Model）分析模型最早是 1994 年 HUGHES 提出的[35]，该模型从企业的角度综合考虑客户一般购买行为。BULT 和 WANSBEEK 对 RFM 的定义为：*R*（Recency）是指用户消费的临近性，与客户重复购买的可能性成反比，通常以用户在观测期内的最近消费时间作为衡量指标；*F*（Frequency）是指用户的消费频率，与客户忠诚度成正比，通常以观测期内用户的消费次数作为衡量指标；*M*（Monetary）是指用户的消费能力，与公司对客户的关注度成正比，通常以观测期内用户的消费总额作为衡量标准[36]。基于 RFM 模型的定义，本文对数字社区用户进行价值划分，帮助社区精准服务于用户。另外鉴于数字社区中用户参与不同活动所获得的奖励额度和奖励物品不同，在测算 *R*、*F*、*M* 值时需通过最大最小归一化方法将数据标准化，以减少测量误差。

用户参与活动 *m* 的近度 $R_m(m=1,2,\dots m)$，R_m 的含义为最近一次参与活动 *m* 的时间，即最后一次参与项目活动距离设定时间的间隔长度。R_m 越小说明数字社区用户越活跃，对数字社区的价值以及贡献就越大。假设实验数据采集的时间点为 *T*，用户的生命周期为 $T_{nm}(n=1,2,\dots,n;\ m=1,2,\dots,m)$，其中，$T_{nm}$ 表示用户参与活动 *m* 的时间点。用户参与活动的近度计算公式如公式（1）所示。

$$R_m=\min(T-T_{nm}) \quad (1)$$

用户参与活动 *m* 的频度 $F_m(m=1,2,\dots m)$，F_m 的含义为顾客一段时间内参与活动 *m* 的次数，参与频率越高代表用户忠诚度越高。假设用户在参与活动 *m* 的各个时间点上的参与次数集合为 $F_{nm}(n=1,2,\dots,n;\ m=1,2,\dots,m)$，其中 f_{nm} 表示用户在时间点 *n* 上参与活动 *m* 的频次。用户参与活动的频度计算公式如公式（2）所示。

$$F_m=\sum_{i=1}^{n} f_{nm} \quad (2)$$

用户参与活动 *m* 的值度 $M_m(m=1,2,\dots m)$，M_m 的含义为用户一段时期内参与活动 *m* 的消费总额。值度越大表示用户对该平台或该活动项目的贡献越大，重要程度越大。本文通过用户参与活动时消耗的游戏货币或积分等作为衡量用户价值贡献和重要性程度。假设用户在各个参与活动的时间点上的消费额度为 M_{nm}(*n*=

$1,2,\dots,n$; $m=1,2,\dots,m$)，其中 M_{nm} 表示用户在参与活动 m 的时间点 n 上的消费额度。则用户参与活动的值度计算公式如公式（3）所示。

$$M_m = \sum_{i=1}^{n} m_{nm} \tag{3}$$

由于每个用户可能同时参与多个活动项目，因此本文测算 R、F、M 值，取每个用户参与的所有活动项目的 R、F、M 对应的平均值。

3.2 用户价值类别标签

基于聚类算法无监督且事先不知道是否被明确分类的特点，本文将“类内高聚合、类间低耦合”作为指导思想。本文将每个用户的 R、F、M 均值，作为用户相似度测量的指标。此外，为了加速收敛，采用 K-Means++ 算法，将未聚类的数据看作在多维空间上的点，采取“欧式距离”作为测量指标，计算每个对象与中心对象的距离，并根据最小距离重新对相应对象进行划分，然后重新计算每个聚类均值直至没有对象再被重新分配给其他类，且聚类中心不再变化。并将误差平方和（SSE）作为度量聚类效果的目标函数，选取 SSE 最小的分类结果作为最终的聚类结果。计算公式如（4）、（5）所示。

欧氏距离计算公式[37]：

$$d(i,j) = \sqrt{(x_{i1}-x_{j1})^2+(x_{i2}-x_{j2})^2+\cdots+(x_{ip}-x_{jp})^2} \tag{4}$$

误差平方和（SSE）计算公式[38]：

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2 \tag{5}$$

3.3 用户–评分矩阵

RFM 是评估和计量用户 CLV 的典型模型，考虑到数字社区用户评分存在失真以及数据稀疏性的缺点。本文借助数字社区用户行为数据量化分析用户 CLV，并以此作为基于用户的协同过滤算法的用户–评分矩阵，用户 CLV 越大，表示用户的满意度越高，评分越高。熵可用来衡量事物出现不确定性的概念[39]，信息熵理论认为，信息是对系统有序状态的度量，而熵是系统无序状态的度量。一般来说，某项指标的信息熵与该项指标所提供的信息量、在综合评价中起的作用以及该项指标的权重成反比。由于 RFM 模型中的 R、F、M 三个变量对用户的 CLV 的贡献不同，本研究借助熵权法计量 3 个变量在影响用户对活动喜爱程度中的比重，将其作为数字社区用户的项目偏好比。最后根据公式（6）得到加权 RFM 值，并以此构建相应的用户项目–评分矩阵。

$$CLV = w_R R_m + w_F F_m + w_M M_m \tag{6}$$

其中，R_m、F_m 和 M_m 分别表示对应活动 m 的 R_m、F_m 和 M_m 指标，w_R、w_F 和 w_M 表示 R_m、F_m 和 M_m 的权重。

3.4 综合相似度计算

余弦相似度是协同过滤推荐算法中衡量用户相似度的一种常用方法。在协同过滤算法中，它通过计算用户或项目之间的余弦相似度来评估用户或项目之间的相似度。因此，本文所构建的 CLV-PM 模型借助余弦相似度衡量用户的相似度。

用户间余弦相似度的计算公式如公式（7）[40]所示。其中，$sim(u,v)$表示用户 u 与用户 v 的综合相似度，分子表示 u 的向量和用户 u' 向量的乘积，分母表示两者模长的乘积。

$$s(u,u') = \frac{\vec{u}\cdot\vec{u'}}{\|\vec{u}\|\times\|\vec{u'}\|} = \frac{\sum_{i\in I_N\cap I_{N'}}(r_{u,i}\times r_{u',i})}{\sqrt{\sum_{i\in I_N\cap I_{N'}} r_{u,i}^{\,2}}\sqrt{\sum_{i\in I_N\cap I_{N'}} r_{u',i}^{\,2}}} \tag{7}$$

3.5 数字社区用户偏好画像

根据公式（7）得到目标用户的 N 个近邻用户之后，依据“目标用户与其相似用户的喜好是相似的”的假设，预测目标用户的偏好。常用的方法是，利用用户相似度和相似用户评分的加权平均值，来获得目标用户的预测评分，按照评分大小降序排序，生成 n 个预测偏好。计算公式如公式（8）所示。

$$\boldsymbol{R_{u,p}} = \frac{\sum_{v\in V}(\boldsymbol{W_{u,v}}.\boldsymbol{R_{v,p}})}{\sum_{v\in V}\boldsymbol{W_{u,v}}} \tag{8}$$

其中，权重 $w_{u,v}$ 是用户 u 和用户 v 的相似度，$R_{v,p}$ 是用户 v 对项目 p 的预测评分。在获得用户 v 对不同项目的预测评分后，选择前 n 个项目生成预测偏好矩阵列表。并据此构建用户偏好画像如图 2 所示。

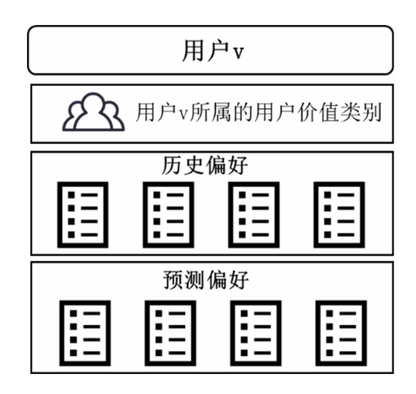


图 2 用户偏好画像

Fig.2 User preference portrait

4 基于 CLV-PM 模型的数字社区用户偏好挖掘研究

4.1 数据源与数据预处理

广西中烟工业有限责任公司通过“微信社群”管理平台对加入平台的用户进行管理和维护，该平台具备个人信息维护、消息推送以及开展营销活动的功能，是一个功能完善且用户累积量较大的数字社区平台。基于此，本文以广西中烟工业有限责任公司“微信社群”的平台数据作为数据源，该数据集包含猜成语、猜歌名和猜码图等 14 个活动的参与情况。具体的活动列表以及各表的数据结构如表 2、表 3 所示。研究涉及该营销平台 2019—2022 年的用户数据，共计 259 268 条。考虑到部分用户数据缺失且不同活动的用户所获得的奖励额度和奖励物品不同，为减少误差，在基于 RFM 模型的计算时需要对不同活动的 R、F、M 值通过最大最小归一化方法使其数据标准化后共得到 38 192 条数据。由于每个用户可能同时参与多个活动项目，因此本文测算的 R、F、M 值取每个用户参与的所有活动项目的 R、F、M 所对应的平均值，最终得到共计 19 362 条数据，数据格式如表 4 所示。

表 2 “微信社群”活动列表

Table 2 List of WeChat community activities

活动名称
猜成语、猜歌名、猜码图、猜谜语、猜诗词、飞行棋大作战、欢乐跳一跳、决胜 21 点、秒杀、趣味大话骰、天天斗地主、歇后语、游戏大厅、众筹

表 3 数据结构

Table 3 Data structure

数据类别	属性编号	属性名称	属性描述
猜成语、猜歌名、猜码图、猜谜语、猜诗词、歇后语	0	UserID	用户的 ID
	1	PicCount	游戏过程中，答对成语（歌名、码图、谜语、诗词、歇后语）的个数
	2	OnePicGold	每个成语（歌名、码图、谜语、诗词、歇后语）奖励的龙币数
	3	CreateTime	创建时间
飞行棋大作战、欢乐跳一跳、决胜 21 点、趣味大话骰、天天斗地主、游戏大厅	0	UserID	用户的 ID
	1	Gold	需要消耗的龙币数
	2	CreateTime	创建时间
秒杀	0	UserID	用户的 ID
	1	GiftNum	秒杀的商品数量
	2	OneUseGold	秒杀一件商品所需的龙币数
	3	CreateTime	创建时间
众筹	0	UserID	用户的 ID
	1	GiftNum	商品数量
	2	TotalValue	选择的商品单个总龙币数
	3	CreateTime	创建时间

表 4　用户的平均 R、F、M 值

Table 4　Average R, F, and M values of users

UserID	R	F	M
1	0.657 75	0.004 53	0.001 13
2	0.272 15	0.241 80	0.028 92
3	0.696 11	0.000 00	0.013 79
4	0.441 25	0.055 96	0.144 98
5	0.867 47	0.007 85	0.001 23
6	0.425 19	0.003 33	0.001 89
…	…	…	…

4.2 数字社区用户价值类别标签

4.2.1　最佳聚类类别数

随着分类数量 k 的增加，误差平方和 SSE 的数值也会变得越来越小，但并非分类数量越多越好。因此，利用“肘部法则”选择“拐点处的 k 值”确定最佳聚类类别数 k。借助 Python 算法不断迭代最终得到如图 3 所示的“手肘图”。由图 3 可知，聚类数从取值为 4 开始，曲线趋于平缓，表明最佳聚类数的取值可能为[4,8]区间内的整数值。

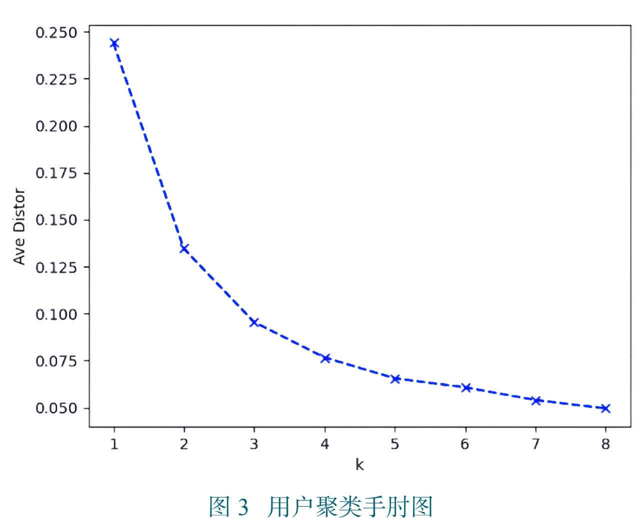


图 3　用户聚类手肘图

Fig.3　User clustering chart

为进一步确定最佳聚类类别数 k，本文借助轮廓系数来确定最终聚类数。轮廓系数得分越高，表示具有定义的聚类模型越好[38]。

轮廓系数的计算公式为：

$$S=\frac{b-a}{max(a,b)} \tag{9}$$

其中，a 表示一个样本与同类中所有其他点之间的平均距离；b 表示样本与下一个最近聚类中所有其他点之间的平均距离。

借助 Python 算法对取值为[4,9]区间内的整数 k 进行多次迭代，最终得到不同 k 取值下的轮廓系数曲线图如图 4 所示，当 k=4 时，轮廓系数最接近于 1，此时模型效果最好，因此，本文的聚类最佳类别数设定为 4 类。

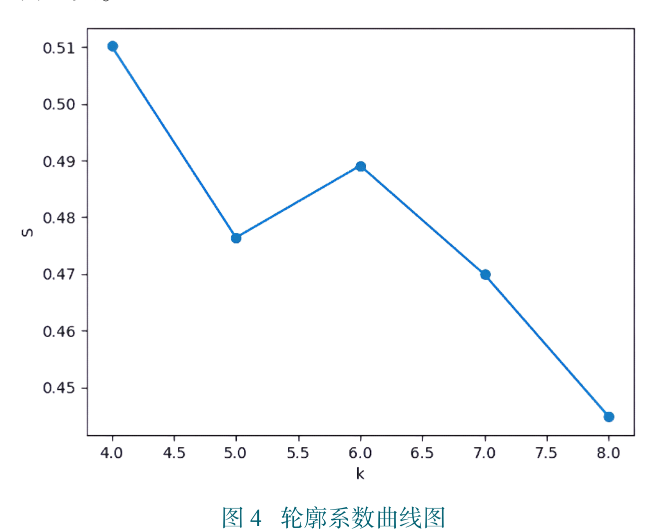


图 4　轮廓系数曲线图

Fig.4　Contour coefficient curve

4.2.2　数字社区用户价值类别划分

借助 Python 编程将属于各个类别的用户分别输出，并计算各类用户的平均值以及总的平均值，得到如表 5 所示的用户分类结果，表中列出了 4 类客户的客户数，平均最近参与时间、参与频率和参与金额以及每一类客户相对于总平均数的 R、F、M 变化情况。

HA 等[41]提出自组织特征映射网络（Self-Organizing Feature Map，SOM）对客户 RFM 指标进行分类，按照用户的价值划分为重要和一般价值客户、重要发展和保持客户、一般发展和保持客户、重要和一般挽留客户共 8 种价值类型。本文参考 HA 的用户分类，结合上述聚类结果特征，得到如下聚类类别，每一个类别的用户都可以被看作是该公司的市场阶段。

（1）重要价值用户（类别 1）：类别 1 的用户参与活动的频度（F）和参与活动的值度（M）较总体平均值大，参与活动的近度（R）较总体平均值较小。说明类别 1 参与活动的次数较多，在活动中的消费额度较

表 5　用户价值类别

Table 5 User value category

类别号	用户数/个	*R*	*F*	*M*	类型
1	6 222	0.112 62	0.039 69	0.021 54	*R*↓*F*↑*M*↑
2	3 131	0.873 35	0.002 91	0.006 77	*R*↑*F*↓*M*↓
3	4 171	0.610 30	0.006 40	0.011 93	*R*↑*F*↓*M*↑
4	5 838	0.352 86	0.014 45	0.014 52	*R*↓*F*↓*M*↓
总平均	19 362	0.415 29	0.018 96	0.014 97	

* 注：*R*、*F* 和 *M* 高于平均值时记为↑，低于平均值时记为↓

大，且近期参与社区内的活动，总体来看较为活跃，对社区的贡献和价值较大，故将其定义为重要价值用户。由重要价值用户偏好统计表可知（表 6），猜码图、猜成语以及游戏大厅等活动项目的用户价值和用户积极性均处于较高水平，属于重要价值用户的高价值高积极性的活动；趣味大话骰和飞行棋大作战的用户价值较高、用户积极性较低，属于高价值低积极性的活动；猜歌名的用户积极性较高，但用户价值处于较低的水平，属于低价值高积极性的活动。针对重要价值用户社区理应采取留存和维持为主的运营策略，提高用户满意度，延长用户生命周期。针对高价值高积极性的活动，社区应当认真建设和完善热度较高的项目；针对高价值低积极性的活动，企业需要挖掘和预测该类用户的偏好，根据偏好进行推送；针对低价值高积极性的活动，社区应当挖掘高热度活动的优点，并将其运用于其他活动中，提高用户的消费意愿。

（2）低价值用户（类别 2）：类别 2 的用户参与活动的频度（*F*）和参与活动的值度（*M*）较总体平均值小，参与活动的近度（*R*）较总体平均值大。说明类别 2 的用户参与活动的次数较少，在活动中的消费额度较小，且最近一次参与社区内的活动时间距今久远，总体来看用户的积极性不高，且对社区的贡献和价值较小，故将其定义为低价值用户。由低价值用户偏好统计表可知（表 7），低价值用户在猜码图活动中的用户价值和积极性均处于最高水平，属于高价值高积极性的活动；众筹、趣味大话骰以及飞行棋大作战等活动

表 6　重要价值用户偏好统计

Table 6 Important value user preference statistics

项目	得分均值	参与用户数/个	重要价值用户的用户画像
趣味大话骰	0.118 97	11	
猜码图	0.102 77	708	
飞行棋大作战	0.088 40	28	
欢乐跳一跳	0.083 76	162	
猜成语	0.057 92	2 109	
众筹	0.044 43	146	
游戏大厅	0.040 35	576	
秒杀	0.037 42	1 025	
猜谜语	0.025 80	145	
决胜 21 点	0.021 22	28	
猜诗词	0.020 73	72	
猜歌名	0.017 18	1 003	
歇后语	0.015 72	52	
天天斗地主	0.011 48	157	

表 7　低价值用户偏好统计

Table 7 Low value user preference statistics

项目	得分均值	参与用户数/个	低价值用户的用户画像
猜码图	0.641 67	1 138	决胜21点 趣味大话骰 天天斗地主 众筹 猜歌名 歇后语 秒杀 飞行棋大作战 猜成语 猜谜语 猜诗词 欢乐跳一跳 游戏大厅 猜码图
众筹	0.187 21	79	
趣味大话骰	0.184 13	16	
飞行棋大作战	0.164 66	8	
秒杀	0.140 76	92	
游戏大厅	0.138 67	119	
决胜 21 点	0.118 89	22	
猜谜语	0.107 61	113	
欢乐跳一跳	0.103 96	87	
猜成语	0.103 65	319	
歇后语	0.093 23	78	
猜歌名	0.091 96	582	
猜诗词	0.090 54	102	
天天斗地主	0.046 92	376	

的用户价值较高，但用户的积极性较低，属于高价值低积极性的活动；猜歌名、天天斗地主以及猜成语等活动的用户积极性较高，但用户的价值较低，属于低价值高积极性的活动。针对低价值用户，社区采取“优先级最低”的运营策略。社区可以通过对高积极性活动采取“精准优化，积极投送”的原则，减少低价值用户的流失。

（3）重要挽留用户（类别 3）：类别 3 的用户参与活动的频度（F）较总体平均值小，参与活动的值度（M）和参与活动的近度（R）较总体平均值大。说明类别 3 的用户参与活动的次数较少，最近一次参与社区内活动距今久远，但用户在活动中的消费额度较大，故将其定义为重要挽留用户。重要挽留用户的主动性较弱，但是其对社区的价值贡献较大，后期社区需要重视该类型用户的偏好挖掘，提高用户的积极性，将其转化成为重要价值用户，采取“提高用户粘性，促进用户转化”为主的运营策略。由重要挽留用户偏好统计表可知（表 8），重要挽留用户在猜码图活动中的用户价值和积极性均处于最高水平，属于高价值高积极性的活动；趣味大话骰、飞行棋大作战以及决胜 21 点等活动的用户价值较高，但用户的积极性较低，属于高价值低积极性的活动；猜歌名和猜成语等活动的用户积极性较高，但用户的价值较低，属于低价值高积极性的活动。因此，针对重要挽留用户的高价值高积极性活动，如猜码图、众筹等，社区需要提高推送频率，注重活动的设计，提高重要挽留用户的用户粘性；针对高价值低积极性的活动，社区可以重点挖掘该类用户的偏好，提出针对性的营销策略，同时通过问卷等的调查方式追踪调查，探索导致重要挽留用户积极性低的原因，提升活动的吸引力；针对低价值高积极性的活动，社区可以适当进行推送以维持重要挽留用户的活性。

（4）回流用户（类别 4）：类别 4 用户消费次数（F）和消费金额（M）高，最近消费时间（R）低，是公司的高价值用户。类别 4 的用户参与活动的频度（F）和参与活动的值度（M）较总体平均值小，参与活动的近度（R）较总体平均值小。说明类别 4 的用户参与活动的次数较少，在活动中的消费额度较小，但近期参与社区内的活动，总体来看该类型用户在近段时间有回流的趋势，故将其定义为回流用户。回流用户的主动性较弱，需要社区加强引导和挖掘偏好，采取“召回为主”的运营策略。由回流用户偏好统计表

表 8 重要挽留用户偏好统计

Table 8 Key retention user preference statistics

项目	得分均值	参与用户数/个	重要挽留用户用户画像
猜码图	0.449 30	1 182	
趣味大话骰	0.164 19	3	
众筹	0.144 92	228	
飞行棋大作战	0.132 81	3	
决胜 21 点	0.107 56	8	
秒杀	0.103 35	253	
游戏大厅	0.093 06	161	
欢乐跳一跳	0.081 37	13	
猜成语	0.077 27	824	
猜谜语	0.075 99	135	
歇后语	0.068 45	89	
猜诗词	0.064 88	132	
猜歌名	0.063 85	1 104	
天天斗地主	0.036 66	36	

可知（表 9），回流用户在猜码图活动中的用户价值和积极性均处于较高水平，属于高价值高积极性的活动；众筹、趣味大话骰、飞行棋大作战以及决胜 21 点等活动的用户价值较高，但用户的积极性较低，属于高价值低积极性的活动；猜歌名和猜成语等活动的用户积极性较高，但用户的价值较低，属于低价值高积极性的活动。因此，针对高价值高积极性的活动，社区需要提高推送频率，并分析高价值高积极性活动相较于其他活动的优点，持续开发和建设类似的活动，促进回流用户转化为重要价值用户；针对高价值低积极性的活动，社区可以提高向用户推送的频率，通过设置礼品、积分等奖励提高用户积极性，同时通过问卷等

表 9 回流用户偏好统计

Table 9 Returned user preference statistics

项目	得分均值	参与用户数/个	回流用户的用户画像
猜码图	0.259 10	1 204	
众筹	0.082 96	172	
飞行棋大作战	0.081 93	19	
趣味大话骰	0.076 53	5	
秒杀	0.074 97	126	
决胜 21 点	0.070 86	17	
游戏大厅	0.066 69	220	
猜成语	0.061 18	2 009	
欢乐跳一跳	0.060 43	63	
猜谜语	0.051 82	163	
猜诗词	0.046 42	146	
歇后语	0.041 91	89	
猜歌名	0.035 55	1 469	
天天斗地主	0.022 52	136	

的调查方式追踪调查，探索其用户回流的潜在背后原因；针对低价值高积极性的活动，社区可以重点挖掘该类用户的偏好，提出针对性的营销策略，提高回流用户留存社区的意愿。

4.3　用户–评分矩阵

由于 *R*、*F*、*M* 在对用户价值偏好的影响程度判别上没有固定标准，因此在测算每个活动项目参与用户的 *R*、*F*、*M* 值后，利用熵值法得到这 3 个指标权重值，并通过加权计算得到用户的 RFM 得分，用以评估和计量数字社区用户的 CLV。各个活动项目的 3 个指标权重如表 10 所示。

表 10　活动项目指标权重列表

Table 10 Activity indicators' weights

活动名称	*R*	*F*	*M*
猜码图	0.743 84	0.126 58	0.129 59
猜成语	0.121 17	0.421 10	0.457 73
猜歌名	0.105 81	0.424 07	0.470 12
猜谜语	0.122 27	0.416 72	0.461 01
猜诗词	0.103 23	0.426 86	0.469 91
歇后语	0.105 56	0.440 45	0.453 99
飞行棋大作战	0.123 19	0.483 97	0.392 84
欢乐跳一跳	0.109 12	0.409 11	0.481 77
决胜 21 点	0.121 14	0.401 56	0.477 30
趣味大话骰	0.217 84	0.385 40	0.396 76
天天斗地主	0.048 81	0.411 53	0.539 66
秒杀	0.146 49	0.571 93	0.281 58
众筹	0.185 88	0.661 45	0.152 67
游戏大厅	0.157 09	0.292 19	0.550 72

考虑到“微信社群”的运营数据缺乏用户评分数据，且加权的 RFM 值能够在用户价值和用户时序性方面真实体现用户偏好。因此，本文基于上述的权重指标，根据公式（6）计算不同用户参与不同项目活动的 *R*、*F*、*M* 的加权平均值，作为用户对某个项目活动的综合评分，没有评分记录记为 Null，得到用户–评分矩阵。

4.4　数字社区用户偏好画像

根据用户之间的兴趣相似度，通过基于用户的协同过滤算法，利用用户当前的项目活动参与情况生成用户–预测评分矩阵，根据预测评分预测和挖掘用户偏好。用户–预测评分矩阵如表 11 所示。

表 11　用户–预测评分矩阵

Table 11 User-prediction scoring matrix

UserID	预测偏好项目	预测分数
16254	飞行棋大作战	0.123 134
16254	猜码图	0.112 041
16254	猜成语	0.069 185
16254	歇后语	0.065 411
16254	猜谜语	0.060 205
16254	猜诗词	0.055 367
16254	猜歌名	0.040 119
16254	决胜 21 点	0.040 022
…	…	…

将预测分数按照降序排序，预测分数前 6 的活动作为用户的预测偏好。基于 CLV-PM 模型从用户价值类别，历史偏好以及预测偏好构建数字社区的用户偏好画像，用户编号为 16254 的数字社区用户偏好画像如图 5 所示。用户 16254 为重要价值用户，该用户参与活动的次数较多，在活动中的消费额度较大，且近期参与社区内的活动，总体来看较为活跃，对社区的贡献和价值较大。针对该用户要采取“留存和维持”为主的运营策略。该用户的历史偏好为欢乐跳一跳、秒杀以及趣味大话骰等，在后期可为该用户推送飞行棋大作战、猜码图、猜成语以及歇后语等活动。

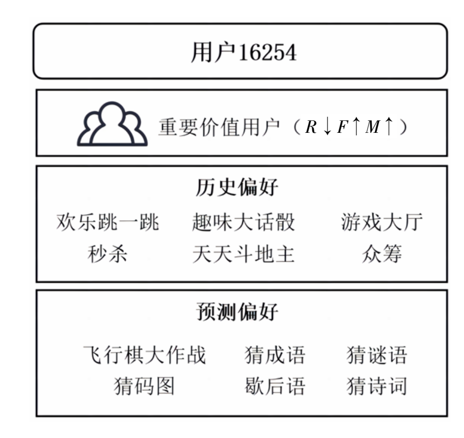


图 5　用户 16254 的用户偏好画像

Fig.5 User preference profile of user 16254

4.5 讨论与分析

基于 CLV-PM 模型对广西中烟工业有限责任公司"微信社群"管理平台用户行为信息开展用户偏好挖掘研究，可以归纳总结为以下结论。

（1）CLV-PM 模型将"微信社群"的用户划分为重要价值用户、低价值用户、重要挽留用户和回流用户 4 个类别。针对重要价值用户，数字社区采取"维持和留存"为主的运营策略；针对低价值用户，数字社区采取"优先级最低"为主的运营策略；针对重要挽留用户，数字社区采取"提高用户粘性，促进用户转化"为主的运营策略；针对回流用户，数字社区采取"召回为主"的运营策略。

（2）针对目标用户绘制用户偏好画像，用户 16254 属于重要价值用户，针对该用户采取"留存和维持"为主的运营策略，该用户的历史偏好为欢乐跳一跳、秒杀等活动，预测偏好为飞行棋大作战、猜码图等活动，为数字社区目标用户的运营和维护提供依据。

5 结　语

随着数字时代的迅猛发展，数字社区被广泛应用于用户管理。针对数字社区用户评分数据失真以及稀疏性的问题以及数字社区用户价值以及时序性的特点，本文提出基于用户 CLV 的数字社区用户偏好挖掘模型 CLV-PM 模型，该模型以用户行为信息作为研究对象，基于 RFM 模型通过 K-means++ 聚类生成用户价值类别标签，将用户的客户生命周期价值作为用户偏好度的衡量指标，并借助协同过滤算法挖掘和预测用户偏好，最后，绘制数字社区用户的用户偏好画像。

（1）在"信息过载"的时代，CLV-PM 模型对数字社区用户的用户偏好挖掘和预测具有重要的实践意义。

（2）CLV-PM 模型以 RFM 模型中的 R、F、M 指标作为聚类依据，在考虑数字社区用户的时序性、差异性以及用户价值的同时，通过 Kmeans++ 对用户进行价值类别分析。在提高目标用户挖掘效率的同时使得数字社区用户价值划分更加明确，减少数字社区的营销成本，提高数字社区的运营绩效，推动数字赋能社区。

（3）CLV-PM 模型以用户行为数据作为数据源，以 RFM 的加权平均值评估和计量用户的 CLV 并将其作为用户对项目的偏好值，基于偏好值挖掘目标用户的预测偏好，实现数字社区用户偏好挖掘研究。该方法在充分考虑了用户价值的情况下有效减少用户评分的失真问题、提高模型对用户偏好的认知，数字社区运营成本得以降低。

（4）本文基于用户 CLV 构建数字社区用户偏好挖掘模型 CLV-PM，为未来客户生命周期价值融入数字社区用户偏好挖掘研究提供新的视角，同时也为用户 CLV 的研究提供新的思路。

参考文献：

[1] GOLDBERG D, NICHOLS D, OKI B M, et al. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61–70.

[2] RAJKUMAR V, KUMAR V. A customer lifetime value framework for customer selection and resource allocation strategy[J]. Journal of marketing, 2004, 68(4): 106–125.

[3] KUMAR V, VENKATESAN R, RAJAN B. Implementing profitability through a customer lifetime value management framework[J]. GfK marketing intelligence review, 2014, 1(2): 32–43.

[4] 李玉婷, 张琅, 姚吉呈. 车险客户生命周期价值研究[J]. 保险研究, 2016(8): 100–114.
LI Y T, ZHANG L, YAO J C. The research on customer lifetime value of auto insurance[J]. Insurance studies, 2016(8): 100–114.

[5] 齐佳音, 马君, 肖丽妍, 等. 考虑客户风险修正的客户终生价值建模[J]. 管理工程学报, 2015, 29(2): 149–159.
QI J Y, MA J, XIAO L Y, et al. Risk-adjusted customer lifetime value measurement[J]. Journal of industrial engineering and engineering management, 2015, 29(2): 149–159.

[6] 朱至文, 周浩东, 孙家溪. 基于客户网络影响力的网络口碑价值度量与预测方法[J]. 统计与决策, 2021, 37(7): 166–169.
ZHU Z W, ZHOU H D, SUN J X. Measurement and prediction

method of online word-of-mouth value based on customer's network influence[J]. Statistics & decision, 2021, 37(7): 166-169.

[7] 陈少霞. 基于价值结构的顾客赢利性测量与管理[J]. 管理工程学报, 2017, 31(2): 109-118.
CHEN S X. The measurement and management of customer profitability: Based on value structure[J]. Journal of industrial engineering and engineering management, 2017, 31(2): 109-118.

[8] 余佳琪, 赵豆豆, 刘蕤. 在线健康社区慢性病患者评论主题情感协同挖掘研究——以甜蜜家园为例[J/OL]. 数据分析与知识发现: 1-20[2023-03-30]. http://kns.cnki.net/kcms/detail/10.1478.G2.20230224.1142.002.html.
YU J Q, ZHAO D D, LIU R. A topic-sentiment collaborative data Mining on the chronic disease patients' reviews in online health community - An evidence from "sweet homeland"[J/OL]. Data analysis and knowledge discovery: 1-20 [2023-03-30]. http://kns.cnki.net/kcms/detail/10.1478.G2.20230224.1142.002.html.

[9] 单思远, 易明. 在线健康社区用户自我信息披露意愿研究[J]. 图书情报工作, 2022, 66(21): 67-77.
SHAN S Y, YI M. Research on self-disclosure intention of online health community users[J]. Library and information service, 2022, 66(21): 67-77.

[10] 肖雪, 李敏, 秦馨怡, 等. 虚拟阅读社区用户互动特征与影响因素[J/OL]. 图书馆论坛: 1-11[2023-03-30]. http://kns.cnki.net/kcms/detail/44.1306.G2.20221116.1232.002.html.
XIAO X, LI M, QIN X Y, et al. Research on users' interaction characteristics and influencing factors in virtual reading community[J/OL]. Library tribune: 1-11 [2023-03-30].http://kns.cnki.net/kcms/detail/44.1306.G2.20221116.1232.002.html.

[11] 普哲缘, 李胜利. 视频评论特征对观众评论行为的影响——以哔哩哔哩为例[J]. 图书情报工作, 2022, 66(20): 130-140.
PU Z Y, LI S L. Influence of video comments characteristics on viewers' commenting behaviors - Taking bilibili as an example[J]. Library and information service, 2022, 66(20): 130-140.

[12] 付少雄, 朱梦蝶, 郑德俊, 等. 基于社会资本理论的在线医疗社区医生知识贡献行为动因研究[J]. 情报资料工作, 2022, 43(3): 67-74.
FU S X, ZHU M D, ZHENG D J, et al. Research on the behavior motivation of doctors' knowledge contribution in online medical community based on social capital theory[J]. Information and documentation services, 2022, 43(3): 67-74.

[13] 赵雪芹, 王青青, 蔡铨. 基于三元交互决定论的在线学术社区动态知识推荐服务模型研究[J]. 农业图书情报学报, 2021, 33(5): 4-13.
ZHAO X Q, WANG Q Q, CAI Q. Dynamic knowledge recommendation service model of online academic community based on ternary interactive determinism[J]. Journal of library and information science in agriculture, 2021, 33(5): 4-13.

[14] 王盼盼, 吴志艳, 罗继锋. 有偿奖励对医生在线健康社区中贡献行为的影响[J]. 系统管理学报, 2022, 31(2): 343-352.
WANG P P, WU Z Y, LUO J F. Effect of monetary incentive on physicians' contribution behavior in online healthcare community[J]. Journal of systems & management, 2022, 31(2): 343-352.

[15] 潘涛涛, 吕英杰. 在线健康社区中基于 SOR 模型的用户参与行为影响因素研究[J]. 情报资料工作, 2022, 43(2): 76-83.
PAN T T, LU Y J. Research on influencing factors of user participation behavior based on SOR model in online health community[J]. Information and documentation services, 2022, 43(2): 76-83.

[16] 赵欣, 李佳倩, 赵琳, 等. 在线社区的知识增殖: 用户行为与用户信任的互惠关系研究[J]. 现代情报, 2020, 40(10): 84-92.
ZHAO X, LI J Q, ZHAO L, et al. Knowledge proliferation in online communities: A research on reciprocal causality between user behavior and user trust[J]. Journal of modern information, 2020, 40(10): 84-92.

[17] 周涛, 王盈颖, 邓胜利. 基于社会资本理论的在线健康社区用户参与行为研究[J]. 信息资源管理学报, 2020, 10(2): 59-67, 129.
ZHOU T, WANG Y Y, DENG S L. Research on online health community users' participation based on social capital theory[J]. Journal of information resources management, 2020, 10(2): 59-67, 129.

[18] 陈星, 张星, 肖泉. 在线健康社区的用户持续知识分享意愿研究——一个集成社会支持与承诺—信任理论的模型[J]. 现代情报, 2019, 39(11): 55-68.
CHEN X, ZHANG X, XIAO Q. Understanding continuous knowledge sharing in the online health communities - An integrated model of social support theory and commitment-trust theory [J].

Journal of modern information, 2019, 39(11): 55-68.

[19] 吴江, 刘涛, 刘洋. 在线社区用户画像及自我呈现主题挖掘——以网易云音乐社区为例[J]. 数据分析与知识发现, 2022, 6(7): 56-69.

WU J, LIU T, LIU Y. Mining online user profiles and self-presentations: Case study of NetEase music community[J]. Data analysis and knowledge discovery, 2022, 6(7): 56-69.

[20] 张军, 李新旺, 李鹏. 多维属性融合视角下的在线健康社区关键用户识别研究[J]. 情报科学, 2022, 40(3): 82-90.

ZHANG J, LI X W, LI P. Key user identification of online health community based on multi-dimensional attribute fusion[J]. Information science, 2022, 40(3): 82-90.

[21] 王帅. 突发公共卫生事件情境下在线健康社区用户画像与分群研究[J]. 情报科学, 2022, 40(6): 98-107.

WANG S. Study on user portrait and clustering of online health community in the context of public health emergencies[J]. Information science, 2022, 40(6): 98-107.

[22] 朱益平, 朱怡, 张诚. 情感体验维度下在线健康社区用户参与行为影响因素研究[J]. 农业图书情报学报, 2022, 34(10): 15-18.

ZHU Y P, ZHU Y, ZHANG C. Influencing factors of user participation behavior in online health community under the dimension of emotional experience[J]. Journal of library and information science in agriculture, 2022, 34(10): 15-18.

[23] 钱宇星, 周华阳, 周利琴, 等. 老年在线社区用户健康信息需求挖掘研究[J]. 现代情报, 2019, 39(6): 59-69.

QIAN Y X, ZHOU H Y, ZHOU L Q, et al. Mining users' health information needs in senior online community[J]. Journal of modern information, 2019, 39(6): 59-69.

[24] 成全, 郑抒琳. 在线健康社区用户信息需求的层级多标签分类研究[J]. 情报理论与实践, 2023, 46(2): 100-108.

CHENG Q, ZHENG S L. Research on hierarchical multi-label classification of user information demand in online health community[J]. Information studies: Theory & application, 2023, 46 (2): 100-108.

[25] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. IEEE transactions on knowledge and data engineering, 2005, 17: 734-749.

[26] 宋凯, 冉从敬. 基于企业画像的高校专利个性化推荐[J]. 图书馆论坛, 2022, 42(9): 123-131.

SONG K, RAN C J. Personalized recommendation of university patents based on enterprise portraits[J]. Library tribune, 2022, 42 (9): 123-131.

[27] 单晓红, 崔凤艳, 刘晓燕. 融合话题多维特征和用户兴趣偏好的微博话题推荐研究[J]. 现代情报, 2022, 42(5): 69-76, 97.

SHAN X H, CUI F Y, LIU X Y. Research on microblog topic recommendation integrating topic multidimensional features and user interest preferences[J]. Journal of modern information, 2022, 42(5): 69-76, 97.

[28] 李亚梅, 秦春秀, 马续补. 基于科研人员情境化主题偏好的科技文献协同推荐研究[J]. 情报理论与实践, 2021, 44(12): 180-189.

LI Y M, QIN C X, MA X B. Research on collaborative recommendation of scientific and technological literature based on researchers' contextual topic preference[J]. Information studies: Theory & application, 2021, 44(12): 180-189.

[29] 李宝. 基于用户画像的高校图书馆个性化资源推荐服务设计[J]. 新世纪图书馆, 2021(4): 68-75.

LI B. Design on personalized resources recommendation service of university library based on user portrait[J]. New century library, 2021(4): 68-75.

[30] 杨辰, 陈晓虹, 王楚涵, 等. 基于用户细粒度属性偏好聚类的推荐策略[J]. 数据分析与知识发现, 2021, 5(10): 94-102.

YANG C, CHEN X H, WANG C H, et al. Recommendation strategy based on users' preferences for fine-grained attributes[J]. Data analysis and knowledge discovery, 2021, 5(10): 94-102.

[31] 李浩, 余雪, 杜旭, 等. 基于学习者时空特征的移动学习资源推荐模型研究[J]. 现代教育技术, 2020, 30(10): 13-19.

LI H, YU X, DU X, et al. Research on the mobile learning resource recommendation model based on learners' temporal and spatial characteristics[J]. Modern educational technology, 2020, 30(10): 13-19.

[32] LI R. Simulation research of university library recommended system based on big data and data mining[C]//Proceedings of the 2015 3rd international conference on machinery, materials and information technology applications, advances in computer science research.

Paris, France: Atlantis Press, 2015: 202–206.

[33] 邢玲, 宋章浩, 马强. 基于混合行为兴趣度的用户兴趣模型[J]. 计算机应用研究, 2016, 33(3): 661–664, 668.
XING L, SONG Z H, MA Q. User interest model based on hybrid behaviors interest rate[J]. Application research of computers, 2016, 33(3): 661–664, 668.

[34] 李建廷, 郭晔, 汤志军. 基于用户浏览行为分析的用户兴趣度计算[J]. 计算机工程与设计, 2012, 33(3): 968–972.
LI J T, GUO Y, TANG Z J. User interest degree calculating based on analysis users' browsing behaviors [J]. Computer engineering and design, 2012, 33(3): 968–972.

[35] JACKSON R, WANG P. Strategic database marketing [M]. Lincolnwood, Ill, USA: NTC Business Books, 1994.

[36] BULT J R, WANSBEEK T. Optimal selection for direct mail [J]. Marketing science, 1995, 14(4): 378–394.

[37] SHIH F Y, WU Y T. Three–dimensional euclidean distance transformation and its application to shortest path planning[J]. Pattern recognition, 2004, 37(1): 79–92.

[38] 吴广建, 章剑林, 袁丁. 基于 K–means 的手肘法自动获取 K 值方法研究[J]. 软件, 2019, 40(5): 167–170.
WU G J, ZHANG J L, YUAN D. Automatically obtaining K value based on K–means elbow method[J]. Computer engineering & software, 2019, 40(5): 167–170.

[39] 罗赟骞, 夏靖波, 陈天平. 网络性能评估中客观权重确定方法比较[J]. 计算机应用, 2009, 29(10): 2624–2626, 2631.
LUO Y Q, XIA J B, CHEN T P. Comparison of objective weight determination methods in network performance evaluation[J]. Journal of computer applications, 2009, 29(10): 2624–2626, 2631.

[40] 武建新, 张志鸿. 融合用户评分与显隐兴趣相似度的协同过滤推荐算法[J]. 计算机科学, 2021, 48(5): 147–154.
WU J X, ZHANG Z H. Collaborative filtering recommendation algorithm based on user rating and similarity of explicit and implicit interest[J]. Computer science, 2021, 48(5): 147–154.

[41] HA S H, PARK S C. Application of data mining tools to hotel data mart on the Intranet for database marketing[J]. Expert systems with applications, 1998, 15(1): 1–31.

User Preference Mining in Digital Community Based on CLV Preference Mining Model

XIAO Yun[1], XU Huanhuan[1], XIAO Yayuan[1], ZHAO Youlin[2,3*], PANG Hangyuan[3]
(1. Guangxi China Tobacco Industry Co., Ltd., Nanning 530001; 2. School of Information Management, Nanjing University, Nanjing 210023; 3. Business School of Hohai University, Nanjing 211100)

Abstract: [Purpose/Significance] Digital communities have become a way for enterprises to manage users efficiently. The existing research on digital community rarely considers the importance of user behavior information and user's customer life cycle value to the mining of user preferences in digital community. This research aims to give full play to the digital community's characteristics such as intuitive, convenient, interesting, and interactive properties so that the research results can benefit every user in their use of the digital community and every enterprise in their user management. [Method/Process] Aiming at the user groups in digital community, this paper proposes a preference mining model ClV-Preference mining (CLV-PM) based on Customer Lifetime Value (CLV). First, in order to

reflect the real preferences of users, the three indicators of the RFM model are used to quantify user behavior information, and the group characteristics of users are mined through K-mean ++ algorithm to generate user value category labels. Second, in order to consider the timeliness and difference of users and enhance the model's cognition of preferences, this paper uses the entropy weight method to solve the indicator weights of each activity, obtains user CLV to construct user-project scoring matrix, and uses the collaborative filtering algorithm to predict user preferences. Finally, based on the user value category, user historical preference and user forecast preference, the user preference profile of target users in digital community is generated, and feasible suggestions are put forward for the operation and maintenance of target users according to the user preference profile. [Results/Conclusions] The user dataset of the "Wechat community" management platform can be divided into four user value categories: important value users, low value users, returned users and important retention users. Target users 16254 are important value users, and the operation strategy of "retention and maintenance" is adopted. The historical preferences are happy hop, sec-kill and other activities; the prediction preference is flying chess battle, guessing code map and other activities; the target user preference sketch provides the basis for the operation and maintenance of users in the digital community. In terms of data source, the CLV-PM model proposed in this paper directly reflects user preferences based on user behavior information and reduces the problem of score distortion. To provide a new perspective for the research of user behavior in digital community, the construction of user-project scoring matrix based on user CLV fully considers the user value of digital community and provides a new direction for the extension and application of CLV. However, due to limited research space, this paper did not conduct model evaluation research on the proposed model, which can be further discussed in subsequent studies.

Keywords: CLV-PM; collaborative filtering; digital community; user preference; information behavior